Article:

Henderson, Lisa M orcid.org/0000-0003-3635-2481, van Rijn, Elaine, James, Emma et al. (3 more authors) (2021) Timing storytime to maximize children's ability to retain new vocabulary. Journal of Experimental Child Psychology. 105207. ISSN 0022-0965

https://doi.org/10.1016/j.jecp.2021.105207

Journal of Experimental Child Psychology 210 (2021) 105207

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp

Timing storytime to maximize children's ability to retain new vocabulary

Lisa M. Henderson *, Elaine van Rijn, Emma James, Sarah Walker, Victoria C.P. Knowland, M. Gareth Gaskell

Department of Psychology, University of York, York YO10 5DD, UK

ARTICLE INFO

Article history:
Received 2 March 2021
Revised 22 April 2021
Available online 20 June 2021

Keywords:
Word learning
Storybook
Memory consolidation
Sleep
Retrieval practice
Testing

ABSTRACT

Shared storybook reading is a key aid to vocabulary acquisition during childhood. However, word learning research has tended to use unnaturalistic (explicit) training regimes. Using a storybook paradigm, we examined whether children (particularly those with weaker vocabularies) are more likely to retain new words if they learn them closer to sleep. Parents read their children (5- to 7-year-olds; N = 237) an alien adventure story that contained 12 novel words with illustrations at one of two training times: at bedtime or 3–5 h before bedtime. Using online tasks, parents tested their children's ability to recall the new words (production) and associate them with pictures (comprehension) immediately after hearing the story and again the following morning. As hypothesized, we replicated two findings. First, children showed overnight improvements in their ability to produce and comprehend new words when tested again the next day. Second, children with better existing vocabulary knowledge showed larger overnight gains in new word comprehension. Counter to expectations, overnight gains in comprehension were larger if the story was read 3–5 h before bedtime rather than at bedtime. These ecologically valid findings are consistent with theories that characterize word learning as a prolonged process supported by mechanisms such as consolidation and retrieval practice, with existing vocabulary knowledge acting as an important source of variability in retention. The findings provide preliminary evidence that encountering new words in stories later in the day (but not too close to sleep) may

* Corresponding author.
E-mail address: lisa-marie.henderson@york.ac.uk (L.M. Henderson).

https://doi.org/10.1016/j.jecp.2021.105207

help to harness vocabulary growth and may be more beneficial than leaving shared storybook reading just for bedtime.

Introduction

It is well established that listening to and reading fiction is important for literacy and language development (Nation, 2017), including the expansion of vocabulary knowledge (Mol, Bus, de Jong, & Smeets, 2008; Montag, Jones, & Smith, 2015), perhaps owing to the rich source of linguistic input that storybooks can provide. Yet, our mechanistic understanding of the word learning process is very much dominated by research studies that use unnaturalistic training paradigms. Such studies have been highly informative, for example, in revealing word learning to be a prolonged process that is bolstered by sleep-based consolidation (see James, Gaskell, Weighall, & Henderson, 2017). However, the use of naturalistic paradigms is vital to advancing theories of word learning and ultimately discovering how we might optimize the process of word learning over the lifespan. Here, we examined how 5- to 7-year-old children learn and retain new words as they listen to their parents reading a story in their own homes and whether such storybook encounters are particularly beneficial if they occur closer to bedtime.

Learning and retaining new words

Many studies now suggest that word learning is a drawn-out process, with the long-term retention of a new word benefitting from repetition (e.g., Horst, 2013; McMurray, Horst, & Samuelson, 2012), retrieval practice (e.g., Goossens, Camp, Verkoeijen, & Tabbers, 2014; Goossens, Camp, Verkoeijen, Tabbers, & Zwaan, 2014; Karpicke & Roediger, 2007; Hulme & Rodd, 2021; van den Broek, Takashima, Segers, Fernández, & Verhoeven, 2013; van den Broek, Takashima, Segers, & Verhoeven, 2018), and reactivation during sleep (see James et al., 2017, for a review). For example, according to a complementary learning systems (CLS) account (McClelland, McNaughton, & O'Reilly, 1995) as applied to word learning (Davis & Gaskell, 2009), encountering a new word rapidly leads to the formation of a short-term hippocampal memory trace. Over time (and particularly during sleep), a longer-term neocortical representation is strengthened within the lexical network. Evidence for this domain-general learning theory (see Chauvette, Seigneur, & Timofeev, 2012; Rasch & Born, 2013) applies to word learning across the lifespan (James et al., 2017).

Studies examining the mechanisms of spoken and orthographic word learning often employ training paradigms in which novel words are encountered in explicit tasks. During this kind of explicit training, participants are required to repeat words aloud and/or write them down and to make phonological, orthographic, or semantic decisions about them (e.g., Bowers, Davis, & Hanley, 2005; Gaskell & Dumay, 2003; Henderson, Weighall, Brown, & Gaskell, 2013a; Wang et al., 2017). Consistent with the CLS account, such studies of spoken word learning have shown that adults and children tend to show improvements in their ability to recall and recognize novel nonsense words after periods of sleep relative to wake (Dumay & Gaskell, 2007; Henderson, Weighall, Brown, & Gaskell, 2012; James, Gaskell, & Henderson, 2020; Smith et al., 2018). Sleep-associated consolidation benefits in performance are also evident when children learn real words via explicit training regimes that incorporate semantic information (Fletcher et al., 2020; James et al., 2020).

There is evidence to suggest that the benefits of sleep outweigh the effect of "repeat testing" (or "retrieval practice") in word learning (Karpicke & Roediger, 2007, 2008; Roediger & Butler, 2011; Rowland, 2014). This refers to the finding that testing memory again after training can improve long-term retention owing to the additional retrieval practice creating further opportunity for learning. Retrieval practice has also been claimed to facilitate the reactivation of new hippocampal memory traces and strengthening of neocortical representations, acting as a "fast route to consolidation"

(Antony, Ferreira, Norman, & Wimber, 2017). Nevertheless, studies of word learning that have adopted 12–12 (or AM–PM) designs, where children and/or adults receive multiple tests that span equivalent periods of wake and sleep, show that improvements in performance are greater when sleep occurs between tests (e.g., Dumay & Gaskell, 2007; Henderson et al., 2012; James et al., 2020). Henderson et al. (2013b) also found that recall of new words was equivalent at a 24-h test regardless of whether children were tested immediately after training or not.

In the current study, we adopted a 0- to 24-h repeat test regime to capture initial learning performance as well as overnight changes in retention. Therefore, we were not concerned with pinning down the precise mechanisms that support any observed retention benefits. Our goal was to better understand variability in the *retention* of new vocabulary following storybook encounters.

Learning and retaining new words from stories

Many studies have shown that exposure to stories fosters vocabulary development in children (e.g., Flack, Field, & Horst, 2018; Justice, Meier, & Walpole, 2005; Montag et al., 2015; Nagy, Anderson, & Herman, 1987; Ricketts, Bishop, Pimperton, & Nation, 2011; Robbins & Ehri, 1994; Sénéchal & Cornell, 1993; Wilkinson & Houston-Price, 2013), with some studies showing that listening to stories (as opposed to reading them) or simultaneously reading and listening to stories is particularly beneficial for word learning in primary school-aged children (Valentini, Ricketts, Pye, & Houston-Price, 2018). Such findings are consistent with evidence that learning words in richer contexts might lead to retention benefits. For example, 1 week after explicit training, Henderson, Weighall, and Gaskell (2013b) observed better phonological recall of new (real) words that were trained with semantic information (in the form of a definition and a picture) as opposed to experiencing the purely phonological and orthographic forms. Whether such benefits are underpinned by richer contexts facilitating links with existing knowledge (e.g., McClelland, 2013) or more simply by enhancing children's motivation to learn and retain the new words (for discussion of this, see Krishnan, Sellars, Wood, Bishop, & Watkins, 2018) remains unclear.

A smaller number of word learning studies have used story learning paradigms to test hypotheses based on the CLS account (Williams & Horst, 2014; Henderson, Devine, Weighall, & Gaskell, 2015; Tamura, Castles, & Nation, 2017). For example, Williams and Horst (2014) used a shared storybook reading task to examine sleep-associated benefits of word learning. Three-year-olds ($N = 48$) listened to the same story multiple times, or heard different stories, followed by either a nap or an equivalent period of time awake. Consistent with the CLS account, children showed sleep-associated gains in word knowledge. Importantly, although there was an initial benefit of hearing the same story multiple times, the nap allowed children who heard different stories to catch up to children who heard the same story repeatedly. Conversely, children who heard different stories and did not nap were not able to catch up to their peers.

Similarly, Henderson et al. (2015) exposed 7- to 10-year-old children ($N = 40$) to novel nonsense words (e.g., daffodat) that were embedded in a story about a trip to an alien zoo that was spoken by a researcher. Children's ability to recall the phonological forms of the novel words (when cued by initial syllables) improved after 24 h, and evidence of lexical integration (captured by lexical competition effects for similar sounding familiar sounds, e.g., daffodil) was not demonstrated immediately after exposure but did emerge 24 h later. Tamura et al. (2017) reported a similar time course of lexical integration when children read unfamiliar written words in fictional texts. As noted above, the memory benefits seen at the 24-h tests in these studies may be a consequence of sleep-based memory reactivation in addition to other mechanisms such as repeat testing. Notably, Hulme and Rodd (2021) recently examined how adults learn new words from reading stories, comparing this with intentional explicit instruction and also examining the role of immediate testing on longer-term retention. They found that story learning (when compared with intentional instruction) led to better performance on cued meaning recall and word-to-meaning matching tasks 24 h after initial exposure. They also found a strong testing benefit for both types of training, but particularly when adults learned words from stories.

Collectively, these studies provide support for the value of storybook encounters for word learning as well as for furthering our understanding of sources of variability in the longer-term retention of new words from stories. However, a major caveat of the developmental literature on storybook word learning is that there is a reliance on researcher-administered stories and lab-style testing of new word knowledge.

Does vocabulary influence longer-term retention?

Another gap in this literature is the examination of variables that predict the extent to which words are retained in long-term memory. One key variable is children's intrinsic vocabulary knowledge. Indeed, Wilkinson and Houston-Price (2013) reported that performance on a standardized test of receptive vocabulary knowledge accounted for more than 20% of variance in comprehension of new words learned from stories both 24 h after exposure and 2 weeks later (see also Penno, Wilkinson, & Moore, 2002). This is consistent with the existence of a *Matthew effect* in vocabulary learning, allowing the "rich to get richer" over time (e.g., Cain & Oakhill, 2011; Coyne, McCoach, Loftus, Zipoli, & Kapp, 2009; Penno et al., 2002). In a meta-analysis, James et al. (2017) reported that existing vocabulary knowledge predicts overnight gains in children's ability to integrate new words (captured by changes in lexical competition effects for existing competitor words) that were learned via explicit training regimes while controlling for initial levels of performance. Therefore, it appears that vocabulary may play a role in the longer-term retention process (including lexical integration) as well as supporting the processes involved in initial acquisition (see also Henderson et al., 2015; Horváth, Myers, Foster, & Plunkett, 2015).

Aligning with this, Henderson and James (2018) exposed children aged 10 and 11 years ($N = 42$) to two sets of eight spoken novel words, with one set embedded in a single story that was presented twice and the other set presented in two different stories. Children with better vocabulary knowledge showed larger overnight gains in their ability to recall the new phonological and semantic word knowledge when novel words were encountered across two stories. However, when novel words were encountered in repetitive stories, vocabulary knowledge exerted no influence on the retention of new word knowledge. This suggests that the extent to which vocabulary knowledge predicts the retention of new words from stories may depend on the task demands. Indeed, James et al. (2021) did not see strong evidence of an association between vocabulary knowledge and changes in memory for new words acquired from stories over time; however, in this study levels of word learning were very low, perhaps leaving little opportunity for vocabulary knowledge to exert an effect.

The mechanisms that underlie an influence of vocabulary on longer-term retention (beyond the influence of vocabulary on the initial learning process) remain somewhat elusive. One proposal is that new memory representations may be more easily integrated into cortical networks if they share similarities with existing representations (Kumaran, Hassabis, & McClelland, 2016; McClelland, 2013; Tse et al., 2007). Based on such accounts, James et al. (2017) hypothesized that richer existing knowledge might support swifter integration of new lexical representations. Broadly speaking, this account is consistent with the above-reported correlations between children's existing vocabulary knowledge and their ability to retain new words. However, it also remains quite possible that children who show bigger retention effects have better general learning abilities that are also reflected in their vocabulary knowledge or that such children are more motivated to learn and retain new words.

Does the time between learning and sleep influence retention benefits?

Another variable recently shown to interact with vocabulary knowledge to predict the extent to which new words are remembered after a delay is the time that elapses between learning and sleep. It is well established that new memories can be subject to interference during wake (e.g., Diekelmann, Wilhelm, & Born, 2009). Subsequently, sleeping sooner after learning has often been claimed to benefit memory in adults (Gais, Lucas, & Born, 2006; McGregor et al., 2013; Payne et al., 2012; Talamini, Nieuwenhuis, Takashima, & Jensen, 2018). However, there is actually little direct evidence to support this, and there is also counterevidence. That is, more time awake after learning has also been shown to be beneficial for long-term memory captured days after learning in adults (Alger, Lau, & Fishbein, 2010; Walker et al., 2019).

However, two recent studies using explicit training regimes have shown that children can benefit from learning new words closer to sleep, particularly if they have poor vocabulary knowledge (James et al., 2020; Walker et al., 2020). In Walker et al. (2020), children aged 8–10 years were trained on orthographic neighbor words (e.g., BANARA derived from BANANA) either in the morning (8–10 a.m.) or in the afternoon (4–6p.m.), and they were tested immediately after training and 24 h later. They were tested using recall, recognition, and also a semantic categorization task that measured lexical competition (i.e., the extent to which human-made/natural response times to BANANA slowed down after having learned BANARA). Larger overnight gains in memory (for both recognition and lexical competition) were observed for children with weaker vocabulary if they learned in the afternoon (i.e., closer to sleep).

James et al. (2020) used a more sensitive within-participants design to tease apart the effects of time versus sleep, where children learned new words in the morning (~8:30–10:00 a.m.; mean ~9:00 a.m.) and in the evening (~6:00–9:30p.m.; mean ~7:45p.m.) before being tested immediately, 12 h, 24 h, and 1 month later. Children learned 12 words on each occasion, all of which were referents for real but rare living things (e.g., hoopoe). The training was again explicit in nature, involving repeating the words aloud and making decisions toward their associations with pictures. At the immediate test, the morning and evening encoding conditions were statistically similar; however, there was an improvement in recall performance 1 month later if they learned in the evening (i.e., closer to sleep) relative to when they learned in the morning. Furthermore, when children learned in the evening, they showed bigger improvements over the 0- to 12-h period than when they learned in the morning. Interestingly, an additional exploratory analysis suggested that these improvements were not dependent on vocabulary; when sleep occurred soon after learning, vocabulary did not help or hinder the consolidation over sleep. However, when children learned in the morning, vocabulary knowledge predicted improvements in recall over sleep. This suggests that if a learning period is followed by a day of wake prior to sleep, children with weaker vocabulary retain less word form knowledge overnight.

There is suggestion that these effects may be developmentally sensitive. Walker et al. (2019) used stimuli and testing tasks similar to those used in Walker et al. (2020) and reported no benefits of learning closer to sleep for adults; in fact, there was some tentative evidence that better recall and recognition of the novel words 1 week after training was associated with *more* time awake between training and sleep (see also Alger et al., 2010). Speculatively, for adults, a delay between learning and sleep might work to enhance declarative memory retention by providing opportunities for additional processing of the new stimuli prior to sleep. It is possible that younger children (and children with weaker language abilities) may be less likely to engage in such wake-based processing and/or that new memory traces are less likely to be tagged for sleep-based consolidation (see James et al., 2020). An alternative (and not mutually exclusive) possibility is that children's new hippocampal memory traces (particularly for children with more impoverished vocabularies) are weaker and more prone to the effects of wake-based interference. Indeed, the hippocampus has a prolonged developmental trajectory that continues until adolescence (Gómez & Edgin, 2016).

The current study

We carried out a real-world test of long-term memory benefits in word learning and the extent to which children's retention can be enhanced by learning closer to sleep when they hear new words in stories read to them by their parents. We recruited slightly younger children (5–7 years) than in Walker et al. (2020) and James et al. (2020), who recruited children aged 8 years and over. If the beneficial effects of learning closer to sleep are due to minimizing wake-based interference, arguably these benefits should be more apparent in a younger population who are at an earlier stage of hippocampal maturity. Taking a novel approach, we asked a large sample of parents to read a story to their 5- to 7-year-old children either 3–5 h before bedtime (i.e., late afternoon) or as a bedtime story. Using online tasks facilitated by parents, we tested the children's ability to recall and recognize the new words both immediately after hearing the story (the "Day 1" test) and the next morning (the "Day 2" test). These tasks were selected to be in keeping with how word learning is often measured in the shared storybook literature (see Flack et al., 2018), with a cued recall task tapping into word production (i.e., children were provided with an initial syllable cue and prompted to recall the whole word) and the recognition task tapping into word comprehension (i.e., children were presented with a spoken word and selected one of four pictures). Previous studies have demonstrated that these tasks can capture sleep-associated improvements following explicit instruction (Henderson et al., 2012; James et al., 2020) and can capture 0- to 24-h improvements after listening to new words in stories (Henderson et al., 2015). We also asked parents to administer a separate online definitions task (on a list of untrained words) to capture their children's vocabulary knowledge (following Henderson et al., 2015; Horváth et al., 2015). We addressed the following preregistered research questions and hypotheses (https://osf.io/bnmza)[1]:

1. Do children aged 5–7 years show increases in their ability to recognize and recall novel nonsense words learned from spoken stories (when tested again) after a period of sleep? We predicted that children would show overnight gains in recognition and recall of novel words encountered in stories (i.e., better recognition and recall of novel words the morning after relative to immediately after hearing the story).
2. Do children show greater levels of retention of novel nonsense words (i.e., overnight improvements in their ability to recognize/recall the novel words) when they learn immediately prior to going to sleep versus 4 h before bedtime? We expected that children would show a "bedtime story effect"—superior overnight gains when they learn immediately prior to sleep relative to when there is a delay between learning and sleep (i.e., as expressed by an interaction between the day and delay conditions).
3. Is this bedtime story effect (i.e., the benefit of learning closer to sleep) influenced by vocabulary knowledge? That is, do children with weaker vocabulary knowledge show more of a benefit of learning closer to sleep? We hypothesized that vocabulary would interact with the bedtime story effect; that is, children with weaker vocabulary knowledge would show a greater benefit of learning closer to sleep (i.e., greater overnight improvement following learning at bedtime relative to in the afternoon) than children with better vocabulary knowledge.

Method

Participants

Families were recruited for this study through pre-established links with schools, social media, and parent sites. Parents were invited to complete a brief online questionnaire to express interest in the study and to check their children's eligibility. A total of 294 eligible typically developing monolingual English-speaking children took part, aged 5–7 years with no known visual, auditory, language, psychiatric, or sleep disorders. However, given that this was an online parent-led study, there was considerable data loss. Of the 294 children, 57 datasets were immediately excluded for reasons such as missing data and noncompliance with the task instructions (e.g., not reading the story or completing the tests at the appropriate times). This left a total sample of 237[2] (n = 111 in the delay condition; n = 126 in the immediate condition). The two groups did not significantly differ in age (delay condition mean = 5.83 years, SD = 0.8; immediate condition mean = 5.79 years, SD = 0.8; p = .736). Each child received a £10 Amazon voucher as a "thank you" for taking part. Prior to recruitment and testing, ethical approval was obtained from the ethics committee, Department of Psychology, University of York.

[1] It should be noted that there are slight changes in the wording of the preregistered hypotheses. These changes were made in response to reviewers' comments, where it was suggested that we make the repeat testing design clearer in our predictions.

[2] We preregistered a desired sample size of 270 (i.e., 135 per group) following the recommendation by Brysbaert and Stevens (2018) of 1600 data points per condition; although these are recommendations for response time (RT) data, they arguably apply when estimating sample sizes for small effects (as anticipated in a parent-led online study). Thus, with 12 novel words to be learned, 135 participants per delay condition was estimated to provide us with the predicted power to detect an effect. However, recruitment to this study was close to completion at the start of the COVID-19 pandemic, at which point the return rate of booklets came to a halt and our ability to print further booklets ceased. Thus, we stopped at a slightly lower than planned sample size for each group.

Materials and procedure

Novel words and story materials

All materials can be found in Appendix A (in article) and Appendix B (in online supplementary material) (see also at https://osf.io/cq7t8/). Each eligible child recruited received a storybook (created by the researchers) through the mail. The story contained 12 novel words taken (or slightly adapted) from the Horst and Hout (2016) NOUN (Novel Object and Unusual Name) database (blinket, coodle, fiffin, gazzer, jefa, koba, manu, pizer, regli, sibu, tannin, and virdex). All were bisyllabic and were deemed to be straightforward for adult readers to pronounce. Each novel word appeared three times in the story. The story was adapted from one used in Henderson et al. (2015), reduced in length and simplified for use in a younger population. The theme of the story was based around a familiar context (i.e., a trip to a zoo) and was designed to be gender neutral. The story took approximately 5 min to read aloud. To facilitate engagement with the story, a number of colored illustrations were created by a research assistant using PowerPoint (as described in James et al., 2021). Each of the novel words was associated with a novel object, also produced in PowerPoint, which appeared in the illustrated scenes each time the word was presented (i.e., three times). The novel object appeared only once on a page, and attempts were made to distribute the novel words across the story as much as possible to avoid them being clustered at the beginning, middle, or end. Storybooks were printed and posted to families and were structured such that only one page of the story was available at a time (following Flack & Horst, 2018).

Procedure and tasks

The procedure is shown in Fig. 1. The study was conducted using Qualtrics software (Qualtrics, Provo, UT, USA) and printed storybooks. Training time was manipulated such that children were randomly allocated to a bedtime condition or a 3- to 5-h delay condition (i.e., alternating group assignment as families signed up), and parents were sent a storybook though the mail. Parents received instructions via e-mail on how to complete the study. They were asked to read this story to their children on a day when their children were not too tired and when it was convenient to retest them 1 h

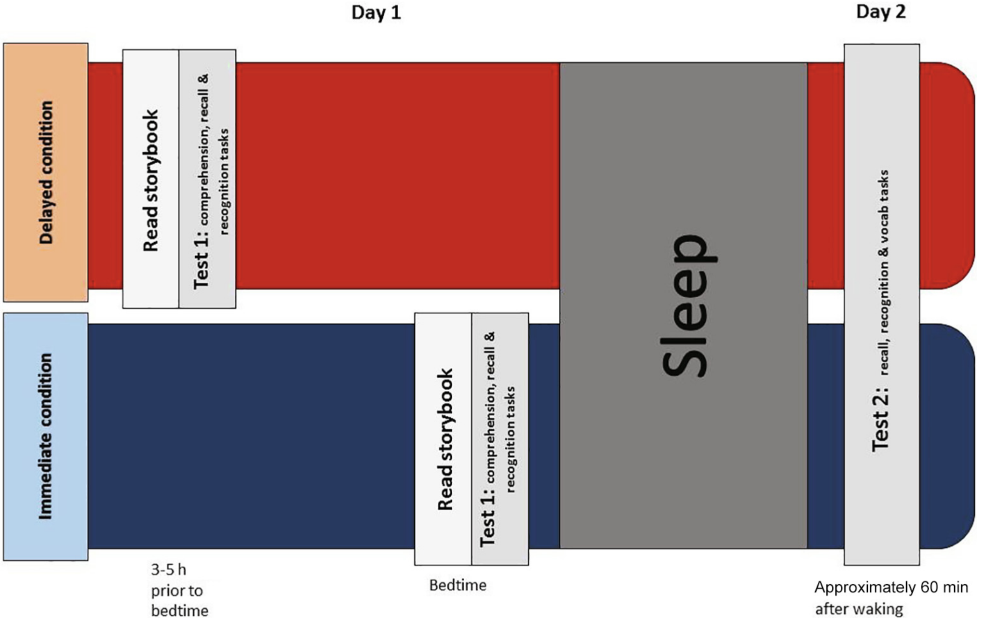


Fig. 1. Experiment procedure.

after they woke up the following morning. In the bedtime condition, parents read the story to their children once they were in bed. In the delay condition, parents read the story to their children 3–5 h before bedtime, with the following instruction given: "It is extremely important that you and your child read this story no earlier than 5 hours before they go to bed and no later than 3 hours before they go to bed. For example, if your child's bedtime is 7p.m., we would estimate that you would read the story together anytime between 2p.m. and 4p.m." In keeping with these instructions, the parent-reported mean start time for the Day 1 test was 15:53 (SD = .03, range = 14:03–17:45) for the Delay group and 19:23 (SD = .03, range = 18:10–21:27) for the Bedtime group. The times when the sessions ended were also available via Qualtrics, verifying that the Delay group ended its Day 1 test at 16:11 on average (SD = .04, range = 13:29–18:07), whereas the Bedtime group ended its Day 1 test at 19:40 on average (SD = .03, range = 18:22–21:47). Importantly, there was no significant difference between the Bedtime group's reported "lights out" bedtime (mean = 19:35, SD = .03) and the Delay group's reported bedtime (mean = 19:43, SD = .03), $p > .05$. Before reading the story, parents were told the following: "It is important that whilst reading to your child, they can see the story too. Please do not read back over any pages, do not repeat any parts of the story, and avoid stopping and answering questions."

Immediately after reading the story (once), parents were instructed to ask their children three comprehension questions to check whether they were attending to the story and then to complete a cued recall task and a recognition task with their children online via Qualtrics using a tablet computer or a phone (see below for details of these tasks). Before beginning these tasks and receiving task-specific instructions, parents were reminded, "For the purposes of the study, please do not read back over any of the storybook pages, repeat any parts of the story to your child, or hint any answers to your child. It is also important that you do not give your child any feedback on their responses." Parents were instructed to administer the cued recall and recognition tasks again the next day (the Day 2 test) roughly 1 h after their children woke up the following morning (Delay group Qualtrics recorded mean end time = 08:19, SD = .04, range = 06:40–10:55; Bedtime group Qualtrics recorded mean end time = 08:17, SD = .04, range = 05:43–11:23). Parents were also asked to rate their children's sleepiness during each session using a customized 10-point scale. At the end of the final testing session, parents were asked to complete the customized measure of expressive vocabulary test with their children.

Comprehension questions

Children were asked three multiple-choice questions as a check to ensure that they were listening to and engaging with the story (see Appendix A). Data for this task were used for exclusion purposes (i.e., to exclude children who scored 0, as described in "Design and analysis" section below). The majority of children answered all questions correctly in each group (Delay group mean = 2.86 out of 3, SD = 0.35; Bedtime group mean = 2.71 out of 3, SD = 0.50, $p > .05$).

Cued recall task

This task captured the number of novel words that children could recall when provided with an initial syllable cue. The instructions given to parents for this task are provided in Appendix A. For each word, the parents/guardians were asked to read aloud to their children a first-syllable cue, which prompted the children to recall the whole word (e.g., blink for blinket, cood for coodle, fiff for fiffin, gazz for gazzer). Parents were asked to write down their children's response. A response was classed as correct if they wrote down the whole word (e.g., blinket), if they wrote down just the part of the word that was missing (e.g., et), or if what they wrote down was phonologically correct (even if orthographically it was misspelled; e.g., blinkit). If they wrote the whole word and the cue was wrong (e.g., blimet), their response was marked as incorrect. All responses were double scored, and any disagreement was resolved between scorers.

Recognition task (word comprehension)

Children were asked to complete a four-alternative forced-choice (4AFC) recognition task. They were asked to decide, out of a choice of four illustrations of novel items from the storybook, which one represented each novel word they had learned. All novel illustrations appeared as targets once and as distractors three times. Responses were scored for accuracy (0 or 1).

Vocabulary task

A customized vocabulary definitions task was completed by each child to capture expressive vocabulary knowledge. The structure of the task and the types of words that were selected (e.g., word classes, categories of words) were based on standardized tests of vocabulary knowledge (e.g., the British Ability Scales–Third Edition [BAS-3] and Wechsler Abbreviated Scale of Intelligence [WASI] Definitions tasks). Children were asked to define the following 10 words in this set order (of increasing difficulty): trousers, tiny, aeroplane, pet, shiver, disappear, enthusiastic, spanner, improvise, and victorious. Caregivers read each word aloud and asked their children to describe the word, typing their children's verbal response. Caregivers were first given a practice round to complete with their children. Two researchers scored the items using a 3-point scoring system developed by the research team (see Appendix A). The maximum possible score on this task therefore was 30. Any disagreements were discussed, checked and resolved.

Design and analysis

A mixed design was used with delay condition (bedtime or delay) as a between-participants variable and day (Day 1 or Day 2) as a within-participants factor. Vocabulary was a continuous predictor.

The data were analyzed using R, with models fitted using the package lme4 (Bates, Maechler, Bolker, & Walker, 2014) and figures created using ggplot2 (Wickham, 2016). Mixed-effects logistic regression models were used to model all outcomes. For each dependent variable (i.e., recall accuracy and recognition accuracy), we fit a model with fixed effects of delay condition (bedtime or 3- to 5-h delay; +.5, −.5), day (Day 1 or Day 2; −.5, +.5), vocabulary, and interactions Day*Delay Condition, Vocabulary*Delay Condition, Vocabulary*Day, and Day*Delay Condition*Vocabulary. Vocabulary scores were centered and scaled prior to model fit. To determine the best fitting fixed-effects structure, a model with a maximal fixed-effects structure and random intercepts only was used. Using a backward selection procedure, each interaction within the fixed-effects structure was removed one at a time, with highest-order interactions explored first. At each stage, the model was compared with each previous model using likelihood ratio tests (LRTs) to determine any model change with a liberal criterion of $p < .20$. Where the removal of a fixed effect did not affect the model (i.e., $p > .20$), the removal of this fixed effect was deemed justifiable. In addition, where the removal of a fixed effect was not justified, all lower-order interactions were retained. This procedure was repeated until all fixed effects were analyzed and the final fixed-effects structure was determined.

Random intercepts and slopes for participant and item effects were also justified using a liberal criterion for model improvement of $p < .20$ via LRTs and added until no further model improvement could be established (i.e., $p < .20$) (Barr, 2013). The intercept (participant or item) that contributed most when compared with the final fixed-effects model was explored first when establishing random slopes, and a forward model selection process was used. Only those main and interaction effects present in the fixed-effects structure were explored. Each main and interaction random slope was added to the intercept one at a time (again using a criterion of $p < .20$). Random slopes for main effects were established first. The model with the lowest *p* value was selected and used as a comparative benchmark when establishing interaction slopes. This was repeated until no further improvement ($p < .20$) could be achieved.

The reported significance tests for each predictor were provided by lmerTest (Kuznetsova, Brockhoff, & Christensen, 2017). Any significant interactions of fixed effects were further tested using pairwise contrasts via the emmeans package in R (Lenth, Singmann, Love, Buerkner, & Herve, 2018).

Data exclusion

Participants were excluded (and replaced during data collection) if they scored 0 on the vocabulary test (no participants) and/or if they scored 0 on the comprehension questions ($n = 1$). Participants were also excluded if they read the story less than 3 h or more than 6 h before bedtime in the delay condition or prior to getting into bed (based on parent report) in the immediate condition ($n = 56$). Dfbetas were calculated to identify any strongly influential cases using the influence.ME package (Nieuwenhuis, Pelzer, & te Grotenhuis, 2012). The dfbetas were standardized, and any participants with *z* scores greater than ± 3.29 were removed from that dataset ($n = 8$ from the recognition task and n = 10 from the cued recall task). We adopted a 3.29 threshold for the removal of influential cases to take a conservative approach and avoid the risk of unnecessary data loss.

Results

As shown in Table 1, recognition performance was good (and well above chance) at both time points, demonstrating that word learning had taken place. Recognition performance was also substantially better than cued recall performance, in keeping with previous studies. It is also important to note that the two groups did not differ on their mean scores on the vocabulary task (delay condition mean = 10.26 out of 30, SD = 3.46; bedtime condition mean = 9.90 out of 30, SD = 3.19), p = .402. The final linear mixed-effects models are presented in Tables 2 and 3.

Hypothesis 1. Do children aged 5–7 years show increases in their ability to recognize and recall novel nonsense words learned from spoken stories when tested again after a period of sleep?

As predicted, collapsing across delay conditions, we observed significant recall and recognition gains the morning after in comparison with immediately after hearing the story (recognition: b = .36, SD = .08, p < .001; cued recall: b = .70, SD = .07, p < .001). Despite a numerical difference in cued recall performance between groups on Day 1 (see Table 1), this difference was not statistically significant, $t(236) = -1.07$, p = .288.

Hypothesis 2. Do children show greater consolidation of novel nonsense words when they learn immediately prior to going to sleep versus 4 h before bedtime?

For cued recall, there was no significant main effect of delay condition (b = .12, SE = .17, p = .50), and counter to Hypothesis 2, there was no significant Delay Condition × Day interaction ($b = -.19$, SE = .14, p = .18) (see Fig. 2). However, for recognition, despite no main effect of delay (b = .21, SE = .14, p = .14),

Table 1
Performance for the recognition and cued recall tasks.

Condition	Day	Recognition *N*	Recognition % correct [*M* (*SD*)]	Cued recall *N*	Cued recall % correct [*M* (*SD*)]
Bedtime	1	122	68.6 (19.8)	121	23.9 (20.9)
	2	122	69.7 (21.7)	121	35.7 (25.1)
Delay	1	107	69.7 (19.1)	106	27.2 (19.1)
	2	107	76.3 (20.9)	106	37.3 (23.5)

Table 2
Predictors of recognition accuracy performance.

Fixed effects	*b*	*SE*	*z*	*p*
Intercept	**1.20**	**.21**	**5.62**	**<.001**
Delay condition	.21	.14	1.48	.14
Day	**.36**	**.08**	**4.67**	**<.001**
Vocabulary	**.42**	**.08**	**5.46**	**<.001**
Delay Condition × Day	**.35**	**.14**	**2.47**	**.01**
Day × Vocabulary	**.24**	**.08**	**3.06**	**.002**
Random effects		Variance		*SD*
Participant: (Intercept)		.90		.95
Participant: Day (slope)		.13		.36
Item: (Intercept)		.48		.70

Note. The final model was based on 5496 observations from 229 participants and 12 items. A further 8 participants were excluded as influential cases. The three-way Delay Condition × Day × Vocabulary interaction was pruned from the model with no significant reduction in model fit, χ^2 = 0.02, p = .89, as was the two-way Delay Condition × Vocabulary interaction, χ^2 = 0.01, p = .92.
Bold text signifies statistically significant main effects or interactions.

Table 3
Predictors of cued recall accuracy performance.

Fixed effects	b	SE	z	p
Intercept	**−1.13**	**.25**	**−4.58**	**<.001**
Delay condition	.12	.17	0.68	.50
Day	**.70**	**.07**	**10.09**	**<.001**
Vocabulary	**.54**	**.12**	**4.52**	**<.001**
Delay Condition × Day	−.19	.14	−1.34	.18
Delay Condition × Vocabulary	−.14	.18	−0.76	.45
Day × Vocabulary	.04	.08	0.50	.62
Delay Condition × Day × Vocabulary	.17	.15	1.15	.25

Random effects	Variance	SD
Participant: (Intercept)	1.36	1.17
Item: (Intercept)	.64	.80
Item: Vocabulary (slope)	.07	.26

Note. The final model was based on 5448 observations from 227 participants and 12 items. A further 10 participants were excluded as influential cases.
Bold text signifies statistically significant main effects or interactions.

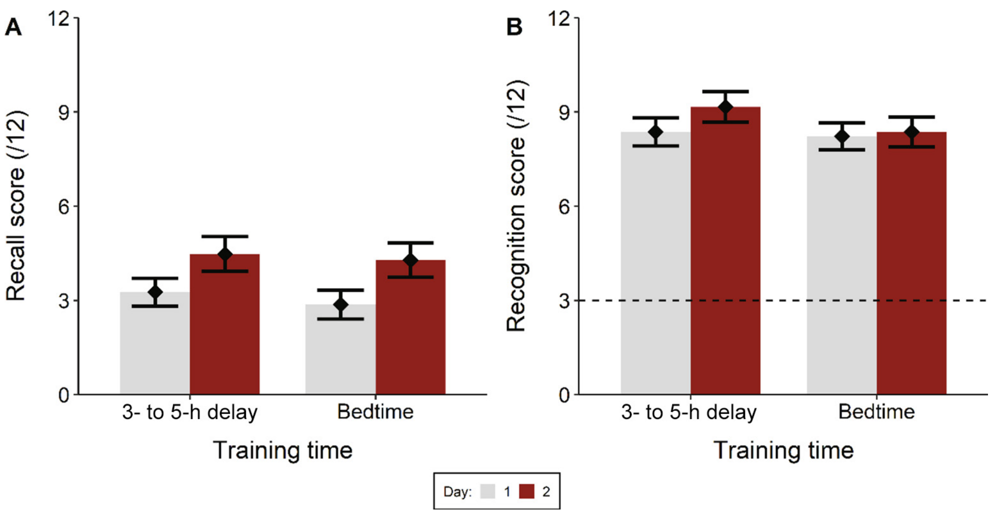


Fig. 2. The Day (1 or 2) × Condition (immediate or delayed) interaction for cued recall (A) and recognition (B). Error bars are 95% confidence intervals. The dashed line marks chance-level performance.

there was a significant Delay Condition × Day interaction (b = .35, SE = .14, p = .01), which was explored further using the R package emmeans (Lenth et al., 2018). For the 3- to 5-h delay condition there was a significant overnight improvement in performance from the Day 1 test to the Day 2 test (b = −.54, SE = .11, p < .001), whereas for the bedtime condition only a trend for an improvement was found (b = −.18, SE = .10, p = .064) (see Fig. 2).

Hypothesis 3. Do children with weaker vocabulary knowledge show more of a benefit of learning closer to sleep?

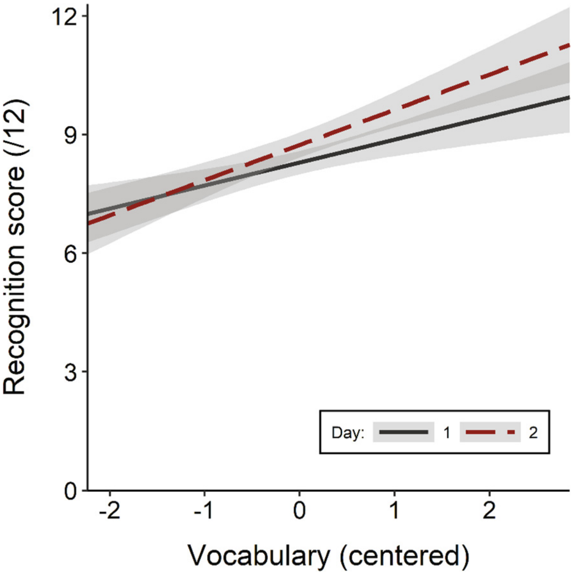


Fig. 3. The Day (1 or 2) × Vocabulary interaction for recognition. Gray bands represent 95% confidence intervals.

Counter to Hypothesis 3, there was no significant Delay Condition × Day × Vocabulary interactions for cued recall (b = .17, SE = .15, p = .25). Removal of the Delay Condition × Day × Vocabulary fixed effect did not significantly affect model fit (p > .20) for the recognition task. Thus, recognition accuracy (akin to recall accuracy) was not influenced by a Delay Condition × Day × Vocabulary interaction.

However, in relation to the effect of vocabulary on learning and retention more generally, collapsing across days, higher accuracy on both recall and recognition measures was positively associated with vocabulary (recognition: b = .42, SE = .08, p < .001; cued recall: b = .54, SE = .12, p < .001), and we did observe a significant Day × Vocabulary interaction for recognition (b = .24, SE = .08, p = .002). As Fig. 3 shows, vocabulary scores were more strongly associated with recognition on Day 2 than on Day 1, with children with better vocabulary showing better recognition on Day 2. This was corroborated using the *emtrends* function within the emmeans package, which indicated that there was a larger simple slope found on Day 2 (trend = .54, SE = .10, confidence interval (CI) = [.35, .72]) in comparison with Day 1 (trend = .30, SE = .07, CI = [.15, .45]).

Discussion

This study took a naturalistic approach to examining the learning and retention of new words from shared storybook reading in a large sample of 239 children aged 5–7 years. Caregivers read their children a story containing novel nonsense words and tested their ability to recall and recognize the novel words straight after listening to the story (the Day 1 test) and again the following morning (the Day 2 test). Central to this study was examining whether the delay between listening to the story and going to sleep had an impact on retention of the new words, with recent studies (e.g., James et al., 2020; Walker et al., 2020) suggesting that a bedtime story may elicit retention benefits particularly for children with weaker vocabulary knowledge. Ultimately, this study aimed to advance our theoretical understanding of the vocabulary acquisition process and also uncover practical implications for how home literacy encounters might be optimized to foster vocabulary growth.

We first tested the hypothesis that children would show overnight gains in recognition and recall of novel words encountered in stories, based on findings from previous studies of word learning from stories (e.g., Henderson et al., 2015; Henderson & James, 2018) and studies of spoken word learning using explicit instruction (e.g., Henderson et al., 2013a, 2013b; James, Gaskell, & Henderson, 2019). This hypothesis was supported. Children showed significantly better recognition and recall of the novel words the morning after listening to the story relative to the previous day. These data are consistent with the CLS account of word learning using an ecologically valid training regime (i.e., that a long-term representation is strengthened over time, particularly sleep, leading to improved memory performance). Although it is very likely that repeat testing (i.e., re-exposure to the items and retrieval practice) partly accounts for these overnight improvements, previous studies have confirmed that such overnight gains in performance are not *purely* a consequence of repeat testing (e.g., Henderson et al., 2012, 2013b; James et al., 2020). Furthermore, sleep parameters measured on the night between training and a Day 2 test (e.g., power within the sleep spindle frequency range and slow oscillation activity) have previously been associated with overnight improvements in new word knowledge (e.g., Fletcher et al., 2020; Smith et al., 2018), suggesting that sleep may be actively supporting these overnight gains.

Also consistent with previous studies of story-based word learning (e.g., Henderson et al., 2015; Henderson & James, 2018), vocabulary predicted word learning ability (collapsing across days) as indexed by both cued recall and recognition tasks. Vocabulary was also a stronger predictor of Day 2 recognition (word comprehension) performance than of Day 1 performance (supported by a Day $\times$ Vocabulary interaction). Thus, children at the higher end of the vocabulary spectrum showed larger overnight improvements in their ability to match the novel words to their picture referents. This finding lends support to the idea that extant vocabulary knowledge is associated with initial vocabulary learning as well as with longer-term retention (see James et al., 2017). On interpreting this finding, however, it is important to note that associations between vocabulary and overnight consolidation of new words have not been consistently observed (see James, 2019), and here we observed this association only for overnight changes in recognition performance and not for changes in cued recall performance. Similar to the current study, James et al. (2021) found this association only for a word learning measure that required mapping new words to referents (similar to our recognition task) and not for recall or recognition of the word forms in isolation. Therefore, existing vocabulary knowledge might play more of a role in the retention of form–meaning mappings and/or when training takes place in richer semantic contexts (i.e., in Henderson et al., 2015, and Henderson & James, 2018). Nevertheless, because the association between vocabulary knowledge and overnight retention benefits in word–meaning mappings did not hold for two further adult experiments in James et al. (2021), further studies are needed to systematically address this hypothesis across development. Alternative mechanistic accounts of this association also need to be examined, including the role of motivation as a possible third variable that could account for the association between vocabulary and the longer-term retention of new words.

There was no support for the hypothesis that children would show a "bedtime story effect" (i.e., superior overnight consolidation when they learned immediately prior to sleep relative to when there was a delay between learning and sleep), nor was there any support for the hypothesis that a bedtime story advantage would be particularly apparent for children with weaker vocabulary (based on James et al., 2020; Walker et al., 2020). In fact, there was more evidence for the reverse pattern for the recognition task. Specifically, larger overnight gains in performance were observed for the group that heard the story 3–5 h before bedtime than for the "bedtime story" group. Importantly, recognition (and recall) performance was statistically equivalent for the two groups at the Day 1 test (which was administered immediately after the story), warding off the potential for group differences in fatigue and/or time of day at the pretest driving this interaction. Nevertheless, future studies adopting a nap paradigm to control time of day effects and match the timing of Day 1 and Day 2 tests would be important to rule this out completely.

The retention advantage in recognition for the Delayed group is, at first glance, seemingly discordant with James et al. (2020) and Walker et al. (2020), who reported that learning closer to sleep led to larger long-term memory benefits for children with lower vocabulary knowledge. However, here children in the bedtime condition heard the story when they were in bed, which contrasts with James et al.'s and Walker et al.'s PM conditions that were typically delivered at least 1 h before bedtime (between 6 and 9:30p.m. on average) in James et al. (2020) and between 4 and 6p.m. in Walker et al.

(2020). Hence, the PM conditions in these previous studies were actually closer to our delay condition in terms of timing before bedtime. Thus, it could be that there is an optimal window to maximize long-term retention in young children that is intermediate: close to sleep, but not too close. In line with this, adult studies have shown some benefits of wake prior to sleep for promoting consolidation and attributed such benefits to homeostatic increases in the need for sleep following learning episodes (e.g., Alger et al., 2010). Of course, we cannot assume from the current data alone that the delayed condition here would be beneficial relative to learning earlier in the day. Particularly owing to the caregiver-led training regime, children (and their caregivers) may also have been more likely to discuss the story and the novel words leading to further opportunities for learning (despite explicit instruction to discourage this); such additional learning opportunities were perhaps less likely in Walker et al. (2020) and James et al. (2020), both of which adopted researcher-led training. Nevertheless, the current data at least suggest that storybook encounters at the end of the school day or after school could help to harness vocabulary growth and may be more beneficial than leaving shared storybook reading only for bedtime.

One question is why the retention advantage seen for the delayed condition was present for recognition performance (i.e., word comprehension) but not for recall (i.e., word production). It is well established that word comprehension precedes word production in the learning process (e.g., Bornstein & Hendricks, 2012; Hamilton, Plunkett, & Schafer, 2000; Huttenlocher, 1974). Indeed, on Day 2 of this study, the children were able to comprehend roughly 73% of the new words encountered in the story on average (comparable to Biemiller & Boote, 2006, and Flack et al., 2018), whereas word production was much lower (~36%). Previous studies examining sleep-dependent consolidation have suggested that task performance is more likely to improve over sleep when initial performance is not too low but also not at ceiling (e.g., Wilhelm, Metzkow-Mészàros, Knapp, & Born, 2012); thus, speculatively, it could be that recognition performance was around this "sweet spot" and thus more sensitive to the benefits of memory reactivation during sleep. Alternatively, and again speculatively, performance on the recognition task may have been less influenced by repeat testing effects, at least relative to the recall task, perhaps leaving more room for variability to be associated with sleep-dependent consolidation mechanisms. This is because the recall task was always administered first in the testing procedure, and therefore Day 2 recall performance benefitted from the repeat exposure to novel words presented in the recognition task on Day 1. In contrast, there was no additional direct exposure to the word–picture pairs during testing (because words were presented with four picture options and no corrective feedback was provided).

Although the use of the novel parent-led approach is a strength of the current study, allowing word learning to be measured under more naturalistic conditions, less control was then exerted over how the story was read (a variable shown to be important for word learning from stories; see Flack et al., 2018) and how the testing tasks were administered. The latter is particularly likely for the cued recall task, where parents were required to read out the initial syllable cues and type their children's responses, meaning that there was room for bias (conscious or not) for parents to support performance, and indeed this might be another reason for why the these data were less sensitive to interactions with condition and vocabulary. Having said that, performance on this task was low and we also checked for influential cases when analyzing the data. This means that atypical responses (e.g., parents who answered for their children or supported their responses) should have been excluded from analysis. We also recruited a large sample size to counteract additional noise in the data. If anything, these caveats make it all the more notable that the basic expected effects were observed, suggesting that this approach has utility for further examinations of word learning in children.

One important extension in future work will be to ascertain whether the effects observed here hold across a diverse array of parent demographics (e.g., varying home literacy environments across socioeconomic spectra) that were not measured here. Previous research into the impact of the home literacy environment on language and literacy development, and how such encounters could be best tailored to maximize language and literacy growth, has relied largely on questionnaires and video-recorded observations (see Hamilton, 2013; Sénéchal & LeFevre, 2002; Sénéchal, LeFevre, Thomas, & Daley, 1998). Similarly, the literature on shared story-based word learning has mostly recruited teachers or researchers, rather than parents, as the readers (Flack et al., 2018). The current study suggests that large-scale parent-led studies using online testing regimes could contribute to understanding how the home literacy environment works to support vocabulary development. This future research direction has potential for societal impact given that high-quality home literacy experiences have been argued to mitigate the risk posed by low socioeconomic status for children's vocabulary development (Payne, Whitehurst, & Angell, 1994). Because children's ability to garner new words from listening to stories is highly variable (Biemiller & Boote, 2006; Mol et al., 2008), further research to explain this variability will be valuable for providing parents and guardians with evidence-based advice on how best to tailor storytime (see also Flack et al., 2018). For example, it will be important to examine whether the influence of vocabulary knowledge on long-term retention holds when children are given more than one opportunity to hear the story (e.g., Henderson & James, 2018) and/or when children are more motivated to learn new words from the story.

In conclusion, this study found that children aged 5–7 years show retention benefits in novel word learning from stories and also that existing vocabulary knowledge is associated with these benefits. These data are consistent with accounts of word learning that specify the importance of mechanisms such as offline consolidation and retrieval practice, using a training and testing regime with higher ecological validity. Counter to our hypotheses, we did not find that learning immediately before sleep is beneficial for retention; rather, we observed that learning roughly 3–5 h before bedtime led to superior recognition of word–picture referents on the day after learning. Nonetheless, these findings are consistent with previous studies showing that learning close (but not too close) to sleep may be beneficial for long-term vocabulary learning. Further studies using naturalistic paradigms will be important for advancing our understanding of how best to "close the vocabulary gap" (see Quigley, 2018) and also for clarifying the impact that a change in reading regime could actually make on children's vocabulary growth. Both of these are crucial issues given that vocabulary is such a strong predictor of educational success as well as employment and mental health outcomes (Armstrong et al., 2017; Snow, Griffin, & Burns, 2005).

Acknowledgments

We express thanks to the brilliant parents and children who made this study possible. We also thank Rhiannon Pearce for creating the storybook illustrations. This work was funded by an Economic and Social Research Council (ESRC) grant (ES/N009924/1) awarded to L.M.H. and M.G.G.

Data availability

Accompanying preregistrations, data, analysis, and materials can be found at https://osf.io/bnmza.

Appendix A

Words and referents

1. Blinket

2. Coodle

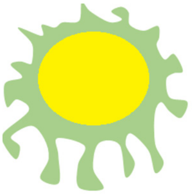
3. Fiffin

4. Gazzer

5. Jefa

6. Koba

7. Manu



8. Pizer

9. Regli

10. Sibu

11. Tannin

12. Virdex

Tasks

Comprehension questions

Now say to your child: "In the story, where did the family go on their day out?" Please read out the following options, and ask your child to choose the correct answer. Please click on the answer your child gives and then press the next button...

Cinema
Space Zoo
Shopping
Don't know

Now say to your child: "In the story, what did mum fill the engine with?" Please read out the following options, and ask your child to choose the correct answer. Please click on the answer your child gives and then press the next button…

Petrol
Glitter
Stardust
Don't know

Now say to your child: "In the story, where did the family have their lunch?" Please read out the following options, and ask your child to choose the correct answer. Please click on the answer your child gives and then press the next button…

Park
Restaurant
At home
Don't know

Cued recall task

Please say to your child: "You might have noticed that there were some funny alien words in that story, that you will have never heard before. In this task I am going to say the first part of each word and I want you to say the full word out loud. For example, if one of the words you learnt was "tomit", I would say "tom" and you would then say "tomit". There were 12 words in total (but we don't expect you to remember them all!)".

Please read out the part-written words below one at a time to your child. Please remind your child to say the FULL word out loud. Please write whatever response your child gives in the text box. If they do not know the answer just leave the text box blank.

This task should take no longer than a couple of minutes (5 mins max). Please complete it in a single sitting.

Blink___
Cood___

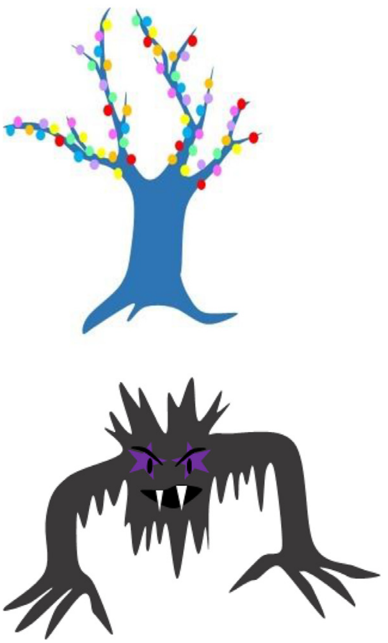
Fiff___
Gazz___
Jef___
Kob___
Man___
Piz___
Reg___
Sib___
Tann___
Vird___

Recognition task (example)

Now say to your child: For the next task I am going to show you four pictures from the story. I will then ask you a question. To answer this question you will need to point to one of the pictures.

Please read out the instruction on each page and click the answer your child points to. After each question press the next button to continue.

1. Point to the jefa



Vocabulary task (example)
For each word you will firstly ask your child if they have ever heard of the word before, and ask them to respond 'yes' or 'no'.

If you select "Yes" you will be required to ask your child to describe the word. Please write their answer in the text box provided. If they cannot describe the word just write 'don't know' in the text box.

If they have not heard the word before select "No" and you will automatically move on to the next word.

Don't worry though, we will provide you with instructions throughout the task!
Remember, when your child is describing what the words mean, we are only looking for roughly a sentence per word (as in the examples). This task should not take you any longer than 10 minutes to complete (approximately a minute per word).

1. Point to the word on the screen and say to your child: *Have you ever heard of the word 'Trousers'?*
   Yes/No

   If the parent selected 'Yes' the following instruction appeared on screen:
   Say to your child: Tell me what 'Trousers' are. Write their answer, exactly as they say it, in the text box provided below. Try not to add any words or change the way they say it:

Scoring system developed for the vocabulary task

Item	1 point	1 point	1 point
trousers	clothes / wear them / put them on	legs / over pants	long / fabric / keep you warm
tiny	small / little	VERY / REALLY small	AND gives an example
aeroplane	flies	vehicle / has wings / has jets/propellers / made of metal	go places / travel far/ go on holiday
Pet	animal (or gives an example) / tap someone or something	lives at your house / you keep it / have it/own it/adopt it	love it / look after it / stroke it / train it / tame it / or gives an example beyond saying it's an animal (note, can't have two points for examples)
shiver	COLD	shaking / moving / example of teeth chattering	also happens when poorly/scared; helps get you warmer
disappear	go / go away / lost / gone / can't see / run away / hidden	invisible / vanish (i.e., a more advanced synonym)	example (e.g., a lost person; no one knows where you are)
enthusiastic	eager / keen / determined / passionate / excited	you really want TO DO SOMETHING / you're really happy ABOUT SOMETHING / feeling good about something	energy / ready to do something / don't give up
spanner	tool / used by engineers /like a wrench	twists/tightens/ turns / screwing / loosen / undo / goes around	bolts OR describes what it looks like / made of metal / fixes things or allows you to build things
improvise	do something	make something up / make something work / guess how to do something / do something your own way	on the spot / creative / do something in a different way
victorious	win / won / winner	really good at something	won a victory or a battle or something challenging / conquered / triumphant / give an example like scoring a goal

Appendix B. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jecp.2021.105207.

References

Alger, S. E., Lau, H. Y., & Fishbein, W. (2010). Delayed onset of a daytime nap facilitates retention of declarative memory. *PLoS One, 5*(8) e12131.

Antony, J. W., Ferreira, C. S., Norman, K. A., & Wimber, M. (2017). Retrieval as a fast route to memory consolidation. *Trends in Cognitive Sciences, 21*, 573–576.

Armstrong, R., Arnott, W., Copland, D. A., McMahon, K., Khan, A., Najman, J. M., & Scott, J. G. (2017). Change in receptive vocabulary from childhood to adulthood: Associated mental health, education and employment outcomes. *International Journal of Language & Communication Disorders, 52*, 561–572.

Barr, D. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in Psychology, 4*. https://doi.org/10.3389/fpsyg.2013.00328.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using eigen and S4. *Journal of Statistical Software, 67*.

Biemiller, A., & Boote, C. (2006). An effective method for building meaning vocabulary in primary grades. *Journal of Educational Psychology, 98*, 44–62.

Bornstein, M. H., & Hendricks, C. (2012). Basic language comprehension and production in >100,000 young children from sixteen developing nations. *Journal of Child Language, 39*, 899–918.

Bowers, J. S., Davis, C. J., & Hanley, D. A. (2005). Interfering neighbours: The impact of novel word learning on the identification of visually similar words. *Cognition, 97*, B45–B54.

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition, 1*(1), 9.

Cain, K., & Oakhill, J. (2011). Matthew effects in young readers: Reading comprehension and reading experience aid vocabulary development. *Journal of Learning Disabilities, 44*, 431–443.

Chauvette, S., Seigneur, J., & Timofeev, I. (2012). Sleep oscillations in the thalamocortical system induce long-term neuronal plasticity. *Neuron, 75*, 1105–1113.

Coyne, M. D., McCoach, D. B., Loftus, S., Zipoli, R., Jr., & Kapp, S. (2009). Direct vocabulary instruction in kindergarten: Teaching for breadth versus depth. *Elementary School Journal, 110*, 1–18.

Davis, M. H., & Gaskell, M. G. (2009). A complementary systems account of word learning: Neural and behavioural evidence. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 364*, 3773–3800.

Diekelmann, S., Wilhelm, I., & Born, J. (2009). The whats and whens of sleep-dependent memory consolidation. *Sleep Medicine Reviews, 13*, 309–321.

Dumay, N., & Gaskell, M. G. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychological Science, 18*, 35–39.

Flack, Z. M., Field, A. P., & Horst, J. S. (2018). The effects of shared storybook reading on word learning: A meta-analysis. *Developmental Psychology, 54*, 1334–1346.

Flack, Z. M., & Horst, J. S. (2018). Two sides to every story: Children learn words better from one storybook page at a time. *Infant and Child Development, 27* e2047.

Fletcher, F. E., Knowland, V., Walker, S., Gaskell, M. G., Norbury, C., & Henderson, L. M. (2020). Atypicalities in sleep and semantic consolidation in autism. *Developmental Science, 23* e12906.

Gais, S., Lucas, B., & Born, J. (2006). Sleep after learning aids memory recall. *Learning & Memory, 13*, 259–262.

Gaskell, M. G., & Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition, 89*, 105–132.

Gómez, R. L., & Edgin, J. O. (2016). The extended trajectory of hippocampal development: Implications for early memory development and disorder. *Developmental Cognitive Neuroscience, 18*, 57–69.

Goossens, N. A. M. C., Camp, G., Verkoeijen, P. P. J. L., & Tabbers, H. K. (2014). The effect of retrieval practice in primary school vocabulary learning. *Applied Cognitive Psychology, 28*, 135–142.

Goossens, N. A. M. C., Camp, G., Verkoeijen, P. P. J. L., Tabbers, H. K., & Zwaan, R. A. (2014). The benefit of retrieval practice over elaborative restudy in primary school vocabulary learning. *Journal of Applied Research in Memory and Cognition, 3*, 177–182.

Hamilton, A., Plunkett, K., & Schafer, G. (2000). Infant vocabulary development assessed with a British communicative development inventory. *Journal of Child Language, 27*, 689–705.

Hamilton, L. (2013). The role of the home literacy environment in the early literacy development of children at family-risk of dyslexia. Ph.D. dissertation, University of York. http://etheses.whiterose.ac.uk/4823.

Henderson, L., Devine, K., Weighall, A., & Gaskell, G. (2015). When the daffodat flew to the intergalactic zoo: Off-line consolidation is critical for word learning from stories. *Developmental Psychology, 51*, 406–417.

Henderson, L. M., & James, E. (2018). Consolidating new words from repetitive versus multiple stories: Prior knowledge matters. *Journal of Experimental Child Psychology, 166*, 465–484.

Henderson, L. M., Weighall, A. R., Brown, H., & Gaskell, M. G. (2012). Consolidation of vocabulary is associated with sleep in children. *Developmental Science, 15*, 674–687.

Henderson, L., Weighall, A., Brown, H., & Gaskell, G. (2013a). Online lexical competition during spoken word recognition and word learning in children and adults. *Child Development, 84*, 1668–1685.

Henderson, L., Weighall, A., & Gaskell, G. (2013b). Learning new vocabulary during childhood: Effects of semantic training on lexical consolidation and integration. *Journal of Experimental Child Psychology, 116*, 572–592.

Horst, J. S. (2013). Context and repetition in word learning. *Frontiers in Psychology, 4*. https://doi.org/10.3389/fpsyg.2013.00149.

Horst, J. S., & Hout, M. C. (2016). The Novel Object and Unusual Name (NOUN) Database: A collection of novel images for use in experimental research. *Behavior Research Methods, 48*(4), 1393–1409. https://doi.org/10.3758/s13428-015-0647-3.

Horváth, K., Myers, K., Foster, R., & Plunkett, K. (2015). Napping facilitates word learning in early lexical development. *Journal of Sleep Research, 24*, 503–509.

Hulme, R. C., & Rodd, J. M. (2021). Learning new words from story reading: The benefit of immediate testing. *PsyArXiv*. https://doi.org/10.31234/osf.io/3r4xd.

Huttenlocher, J. (1974). The origins of language comprehension. In R. Solso (Ed.), *Theories in cognitive psychology: The Loyola Symposium*. Hillsdale, NJ: Lawrence Erlbaum.

James, E. (2019). *Understanding individual differences in learning and consolidating new vocabulary* Ph.D. thesis. University of York.

James, E., Gaskell, M. G., & Henderson, L. M. (2019). Offline consolidation supersedes prior knowledge benefits in children's (but not adults') word learning. *Developmental Science, 22* e12776.

James, E., Gaskell, M. G., & Henderson, L. M. (2020). Sleep-dependent consolidation in children with comprehension and vocabulary weaknesses: It'll be alright on the night? *Journal of Child Psychology and Psychiatry, 61*, 1104–1115.

James, E., Gaskell, M. G., Pearce, R., Korell, C., Dean, C., & Henderson, L. M. (2021). The role of prior lexical knowledge in children's and adults' word learning from stories. *PsyArXiv*. https://doi.org/10.31234/osf.io/vm5ad.

James, E., Gaskell, M. G., Weighall, A. R., & Henderson, L. M. (2017). Consolidation of vocabulary during sleep: The rich get richer?. *Neuroscience and Biobehavioral Reviews, 77*, 1–13.

Justice, L. M., Meier, J., & Walpole, S. (2005). Learning new words from storybooks. *Language, Speech, and Hearing Services in Schools, 36*, 17–32.
Karpicke, J. D., & Roediger, H. L. III, (2007). Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 704–719.
Karpicke, J. D., & Roediger, H. L. III, (2008). The critical importance of retrieval for learning. *Science, 319*, 966–968.
Krishnan, S., Sellars, E., Wood, H., Bishop, D., & Watkins, K. E. (2018). The influence of evaluative right/wrong feedback on phonological and semantic processes in word learning. *Royal Society Open Science, 5*(9). 171496. https://doi.org/10.1098/rsos.171496.
Kumaran, D., Hassabis, D., & McClelland, J. L. (2016). What learning systems do intelligent agents need? Complementary learning systems theory updated. *Trends in Cognitive Sciences, 20*, 512–534.
Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*. https://doi.org/10.18637/jss.v18082.i18613.
Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2018). emmeans: Estimated marginal means, aka least-squares means. *R package, Version, 1*(3), 2.
McClelland, J. L. (2013). Incorporating rapid neocortical learning of new schema-consistent information into complementary learning systems theory. *Journal of Experimental Psychology: General, 142*, 1190–1210.
McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review, 102*, 419–457.
McGregor, K. K., Licandro, U., Arenas, R., Eden, N., Stiles, D., Bean, A., & Walker, E. (2013). Why words are hard for adults with developmental language impairments. *Journal of Speech, Language, and Hearing Research, 56*, 1845–1856.
McMurray, B., Horst, J. S., & Samuelson, L. K. (2012). Word learning as the interaction of online referent selection and slow associative learning. *Psychological Review, 119*, 831–877.
Mol, S. E., Bus, A. G., de Jong, M. T., & Smeets, D. J. H. (2008). Added value of dialogic parent–child book readings: A meta-analysis. *Early Education and Development, 19*, 7–26.
Montag, J. L., Jones, M. N., & Smith, L. B. (2015). The words children hear: Picture books and the statistics for language learning. *Psychological Science, 26*, 1489–1496.
Nagy, W. E., Anderson, R. C., & Herman, P. A. (1987). Learning word meanings from context during normal reading. *American Educational Research Journal, 24*, 237–270.
Nation, K. (2017). Nurturing a lexical legacy: Reading experience is critical for the development of word reading skill. *npj Science of. Learning, 2*, 3.
Nieuwenhuis, R., Pelzer, B., & te Grotenhuis, M. (2012). influence.ME: Tools for detecting influential data in mixed effects models. *The R Journal, 4*(2), 38–47. https://journal.r-project.org/archive/2012-2/RJournal_2012-2_Nieuwenhuis~et~al.pdf.
Payne, J. D., Tucker, M. A., Ellenbogen, J. M., Wamsley, E. J., Walker, M. P., Schacter, D. L., & Stickgold, R. (2012). Memory for semantically related and unrelated declarative information: The benefit of sleep, the cost of wake. *PLoS One, 7*(3) e33079.
Payne, A. C., Whitehurst, G. J., & Angell, A. L. (1994). The role of home literacy environment in the development of language ability in preschool children from low-income families. *Early Childhood Research Quarterly, 9*, 427–440.
Penno, J. F., Wilkinson, I. A. G., & Moore, D. W. (2002). Vocabulary acquisition from teacher explanation and repeated listening to stories: Do they overcome the Matthew effect? *Journal of Educational Psychology, 94*, 23–33.
Quigley, A. (2018). *Closing the vocabulary gap*. Abingdon, UK: Routledge.
Rasch, B., & Born, J. (2013). About sleep's role in memory. *Physiological Reviews, 93*, 681–766.
Ricketts, J., Bishop, D. V., Pimperton, H., & Nation, K. (2011). The role of self-teaching in learning orthographic and semantic aspects of new words. *Scientific Studies of Reading, 15*, 47–70.
Robbins, C., & Ehri, L. C. (1994). Reading storybooks to kindergartners helps them learn new vocabulary words. *Journal of Educational Psychology, 86*, 54–64.
Roediger, H. L., III, & Butler, A. C. (2011). The critical role of retrieval practice in long term retention. *Trends in Cognitive Sciences, 15*, 20–27.
Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin, 140*, 1432–1463.
Sénéchal, M., & Cornell, E. H. (1993). Vocabulary acquisition through shared reading experiences. *Reading Research Quarterly, 28*, 360–374.
Sénéchal, M., & LeFevre, J. A. (2002). Parental involvement in the development of children's reading skill: A five-year longitudinal study. *Child Development, 73*, 445–460.
Sénéchal, M., LeFevre, J. A., Thomas, E. M., & Daley, K. E. (1998). Differential effects of 13 home literacy experiences on the development of oral and written language. *Reading Research Quarterly, 33*, 96–116.
Smith, F. R. H., Gaskell, M. G., Weighall, A. R., Warmington, M., Reid, A. M., & Henderson, L. M. (2018). Consolidation of vocabulary is associated with sleep in typically developing children, but not in children with dyslexia. *Developmental Science, 21* e12639.
Snow, C. E., Griffin, P., & Burns, M. S. (2005). *Knowledge to support the teaching of reading: Preparing teachers for a changing world*. San Francisco: Jossey-Bass.
Talamini, L. M., Nieuwenhuis, I. L., Takashima, A., & Jensen, O. (2008). Sleep directly following learning benefits consolidation of spatial associative memory. *Learning & Memory, 15*, 233–237.
Tamura, N., Castles, A., & Nation, K. (2017). Orthographic learning, fast and slow: Lexical competition effects reveal the time course of word learning in developing readers. *Cognition, 163*, 93–102.
Tse, D., Langston, R. F., Kakeyama, M., Bethus, I., Spooner, P. A., Wood, E. R., ... Morris, R. G. (2007). Schemas and memory consolidation. *Science, 316*, 76–82.
Valentini, A., Ricketts, J., Pye, R. E., & Houston-Price, C. (2018). Listening while reading promotes word learning from stories. *Journal of Experimental Child Psychology, 167*, 10–31.
van den Broek, G. S. E., Takashima, A., Segers, E., Fernández, G., & Verhoeven, L. (2013). Neural correlates of testing effects in vocabulary learning. *NeuroImage, 78*, 94–102.

van den Broek, G. S. E., Takashima, A., Segers, E., & Verhoeven, L. (2018). Contextual richness and word learning: Context enhances comprehension but retrieval enhances retention. *Language Learning, 68*, 546–585.
Walker, S., Gaskell, M. G., Knowland, V. C. P., Fletcher, F. E., Cairney, S. A., & Henderson, L. M. (2020). Growing up with interfering neighbours: The influence of time of learning and vocabulary knowledge on written word learning in children. *Royal Society Open Science, 7* 191597.
Walker, S., Henderson, L. M., Fletcher, F. E., Knowland, V. C. P., Cairney, S. A., & Gaskell, M. G. (2019). Learning to live with interfering neighbours: The influence of time of learning and level of encoding on word learning. *Royal Society Open Science, 6* 181842.
Wang, H. C., Savage, G., Gaskell, M. G., Paulin, T., Robidoux, S., & Castles, A. (2017). Bedding down new words: Sleep promotes the emergence of lexical competition in visual word recognition. *Psychonomic Bulletin & Review, 24*, 1186–1193.
Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. New York: Springer-Verlag.
Wilhelm, I., Metzkow-Mészàros, M., Knapp, S., & Born, J. (2012). Sleep-dependent consolidation of procedural motor memories in children and adults: The pre-sleep level of performance matters. *Developmental Science, 15*, 506–515.
Wilkinson, K. S., ... Houston-Price, C. (2013). Once upon a time, there was a pulchritudinous princess: The role of word definitions and multiple story contexts in children's learning of difficult vocabulary. *Applied Psycholinguistics, 34*, 591–613.
Williams, S. E., & Horst, J. S. (2014). Goodnight book: Sleep consolidation improves word learning via storybooks. *Frontiers in Psychology, 5*. https://doi.org/10.3389/fpsyg.2014.00184.